

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-325207

(43)Date of publication of application : 22.11.2001

(51)Int.Cl.

G06F 13/14

G06F 3/06

G06F 12/08

G06F 15/16

H04L 12/56

(21)Application number : 2000-144376

(71)Applicant : HITACHI LTD

(22)Date of filing : 17.05.2000

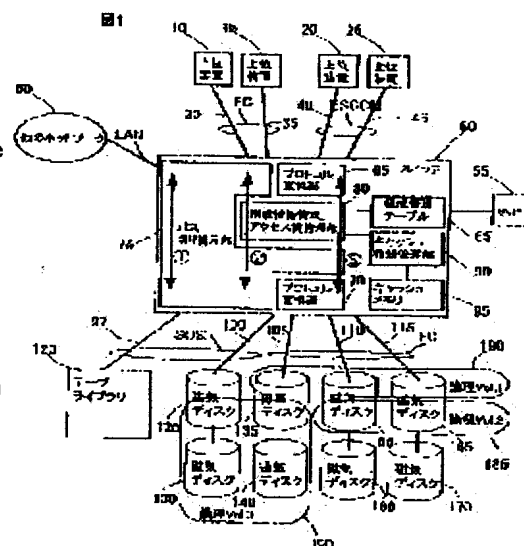
(72)Inventor : TANAKA ATSUSHI

(54) SWITCH WITH BUILT-IN CACHE, COMPUTER SYSTEM AND SWITCH CONTROL METHOD FOR SWITCH WITH BUILT-IN CACHE

(57)Abstract:

PROBLEM TO BE SOLVED: To improve transaction performance by incorporating a cache memory inside a switch in the case of connecting a host device, a cache and a magnetic disk by the switch.

SOLUTION: When an access request to a storage device is sent from the host device to this switch 60, a constitution information management and access right management part 80 inside the switch 60 reads data from the cache memory 95 and sends them through a path to the host device in the case that request data are stored in the cache memory 95 and executes a path switching processing and transfers a transaction through the path to logic Vol, an individual magnetic disk and a tape library, etc., in the case that the request data are not stored in the cache memory 95 by referring to a constitution information table storing the management information of the storage device connected to the switch 60 and the management information of the cache memory 95.



LEGAL STATUS

[Date of request for examination]

27.01.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-325207

(P2001-325207A)

(43) 公開日 平成13年11月22日 (2001. 11. 22)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 13/14	3 1 0	G 0 6 F 13/14	3 1 0 F 5 B 0 0 5
3/06	3 0 1	3/06	3 0 1 B 5 B 0 1 4
12/08		12/08	B 5 B 0 4 5
			H 5 B 0 6 5
			P 5 K 0 3 0

審査請求 未請求 請求項の数 6 O L (全 16 頁) 最終頁に続く

(21) 出願番号 特願2000-144376(P2000-144376)

(22) 出願日 平成12年5月17日 (2000. 5. 17)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 田中 淳

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74) 代理人 100099298

弁理士 伊藤 修 (外1名)

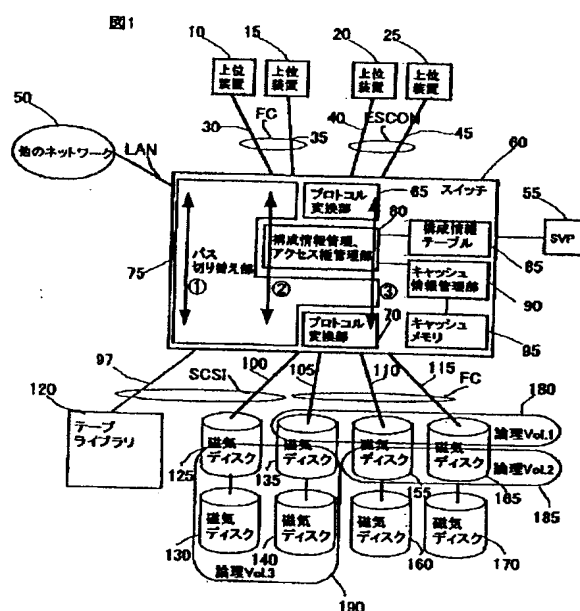
最終頁に続く

(54) 【発明の名称】 キャッシュ内蔵型スイッチおよび計算機システムおよびキャッシュ内蔵型スイッチのスイッチ制御方法

(57) 【要約】

【課題】 上位装置、キャッシュ、磁気ディスクをスイッチで接続する場合、スイッチ内にキャッシュメモリを内蔵し、トランザクション性能を向上する。

【解決手段】 スイッチ60に上位装置から記憶装置へのアクセス要求が送られたとき、スイッチ60内の構成情報管理、アクセス権管理部80は、スイッチ60に接続された記憶装置の管理情報およびキャッシュメモリ95の管理情報を記憶している構成情報テーブルを参照することにより、要求データがキャッシュメモリ95に記憶されている場合には、キャッシュメモリ95からデータを読み取り、バスを介して上位装置に送り、要求データがキャッシュメモリ95に記憶されていない場合にはバス切り替え処理を実行して、トランザクションを論理Vol.1、個別の磁気ディスク、テープライブラリ等にバスを介して転送する。



【特許請求の範囲】

【請求項1】 複数の上位装置と、磁気ディスク、テープライブラリ等の複数の記憶装置にバスで接続され、前記上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチであって、
前記記憶装置の構成情報を記憶する手段と、該構成情報の管理及びアクセス権の管理を行う構成情報管理・アクセス権管理手段と、キャッシュメモリと、該キャッシュメモリを管理するキャッシュ管理手段と、前記バスの切り替えを行うバス切り替え手段を備えることを特徴とするキャッシュ内蔵型スイッチ。

【請求項2】 請求項1記載のキャッシュ内蔵型スイッチにおいて、

該キャッシュ内蔵型スイッチで受信するフレームを該キャッシュ内蔵型スイッチにおけるプロトコルに変換し、該キャッシュ内蔵型スイッチ送信から送信するフレームを外部の接続装置にあったプロトコルに変換するプロトコル変換手段をさらに備えることを特徴とするキャッシュ内蔵型スイッチ。

【請求項3】 複数の上位装置と、磁気ディスク、テープライブラリ等の複数の記憶装置と、該上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチを備える計算機システムであって、
前記キャッシュ内蔵型スイッチは、前記記憶装置の構成情報を記憶する手段と、該構成情報の管理及びアクセス権の管理を行う構成情報管理・アクセス権管理手段と、キャッシュメモリと、該キャッシュメモリを管理するキャッシュ管理手段と、前記バスの切り替えを行うバス切り替え手段を備えることを特徴とする計算機システム。

【請求項4】 請求項3記載の計算機システムにおいて、
前記キャッシュ内蔵型スイッチで受信するフレームを該キャッシュ内蔵型スイッチにおけるプロトコルに変換し、該キャッシュ内蔵型スイッチ送信から送信するフレームを外部の接続装置にあったプロトコルに変換するプロトコル変換手段をさらに備えることを特徴とする計算機システム。

【請求項5】 複数の上位装置と、磁気ディスク、テープライブラリ等の記憶装置にバスで接続され、前記上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチのスイッチ制御方法であって、
受信したフレームから発信元ID、発信先IDを取得し、前記キャッシュ内蔵型スイッチ内に設けられた構成情報テーブルを参照して発信元IDが発信先IDに対するアクセス権を有する場合、該フレーム内の情報が論理V_o1. アクセスを指定しているか否かの判定をし、指定してない場合はバス切り替え処理を行い、該フレームをフレームで指定された記憶装置に送信し、指定してる場合には、前記フレームから論理V_o1. とアドレスを取り出し、前記構成情報テーブルを参照して

該論理V_o1. とアドレスを書き換える実送信IDとアドレスを求め、書き換えを行い、

前記キャッシュ内蔵型スイッチ内に設けられたキャッシュテーブルを前記論理V_o1. とアドレスにより参照してキャッシュヒットの場合、データをアクセスし、該データを発信元に送信するためバス切り替え処理を行い、キャッシュミスの場合、バス切り替え処理を行い、前記書き換えたフレームを実送信先の記憶装置に送信し、記憶装置から返送されたアクセスされたデータをキャッシュメモリに格納し、該データを発信元に送信するためバス切り替え処理を行うことを特徴とするキャッシュ内蔵型スイッチのスイッチ制御方法。

【請求項6】 請求項5記載のキャッシュ内蔵型スイッチのスイッチ制御方法において、

前記受信したフレームのプロトコルが前記キャッシュ内蔵型スイッチにおけるプロトコルと異なる場合には、該受信したフレームのプロトコルを該キャッシュ内蔵型スイッチにおけるプロトコルに変換し、
前記キャッシュ内蔵型スイッチから送信するフレームが送信先の装置のプロトコルと異なる場合には、該キャッシュ内蔵型スイッチから送信するフレームのプロトコルを該送信先の装置のプロトコルに変換することを特徴とするキャッシュ内蔵型スイッチのスイッチ制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、スイッチに係わり、特に複数の記憶装置、CPUなどの構成要素を接続し、高い応答性能、トランザクション数が要求される情報処理装置、記憶装置に有効なスイッチに関する。

【0002】

【従来の技術】 従来の記憶装置では、多数の上位装置、磁気ディスクを接続する場合、トランザクション性能等を向上するために装置間の接続にスイッチを用いることが考えられている。接続のための伝送路およびスイッチについては、具体的な製品の規格としてファイバーチャネルが存在している。またこれを用いて磁気ディスクを多数接続し記憶装置を構成したものとしてSAN(Storage Area Network)が提案されている。このような技術については、日経エレクトロニクス1998年5月4日(n_o.715)に詳しく記載されている。この場合、スイッチは上位装置、キャッシュ、磁気ディスクと接続するためのI/Fを持ち、データの変換を行う。またキャッシュを制御するためにキャッシュ内にキャッシュ制御部およびキャッシュ管理情報を持ち、上位装置からのデータ要求に対してキャッシュ内にデータが存在するか(キャッシュヒット)の判定およびデータの送信を行う。

【0003】 このような記憶装置においてスイッチを用いた技術には、特開平9-198308号に開示された「データ記憶システム」等がある。またキャッシュ内にキャッシュ管理情報を持たず、上位装置内の制御メモ

10

20

30

40

50

リ、または別の共有メモリに管理情報を持ちそこでキャッシュヒットの判定を行い、ヒットであれば直接キャッシュメモリにリードを行う方式も考えられている。このような記憶装置においてスイッチを用いた技術には、特開平11-203201号に開示された「キャッシュメモリの配置およびデータ記憶システム」等がある。

【0004】図2に従来のスイッチ接続型記憶装置の構成図を示す。200、205、210、215は上位装置である。220、225は、上位装置と記憶装置を接続するためのインターフェイス(CHA-PK)である。240、245、250、255はキャッシュである。270、275はキャッシュと磁気ディスク装置を接続するためのインターフェイス(DKA-PK)である。280、285、290、295は磁気ディスクである。230、235、260、265は上位装置200～215や磁気ディスク280～295とキャッシュ240～255を接続するスイッチである。

【0005】

【発明が解決しようとする課題】以上に述べたように、上位装置、キャッシュ、磁気ディスク装置にそれらと互換性を持ったI/Fを持ったスイッチを接続することにより、同時に多数のトランザクションを処理することが可能となり、記憶装置の高性能化が図れる。しかしキャッシュをスイッチに接続できる一つの装置とすることでスイッチとキャッシュの間にコマンド、データの変換が発生し、処理時間を増加させる。またキャッシュ内に求めるデータがあるかどうかの判定を行うためにキャッシュで判定制御を行うか、他のメモリ内にある制御メモリを使って判定する必要がある。どちらの場合もヒット判定にともなうデータ変換が生じ、余計なオーバーヘッドがかかるためトランザクション性能低下につながる。また外部にキャッシュを設置するためには、半導体メモリ装置または、キャッシュ内蔵のディスク制御装置を接続する必要があり、一般のユーザにとっては設備コストを増加させることになる。

【0006】本発明の主たる目的は、上位装置、キャッシュ、磁気ディスクをスイッチで接続する記憶装置においてトランザクション性能を向上するためにスイッチ内にキャッシュメモリおよびキャッシュ管理情報を内蔵することでスイッチ技術を改良することにある。本発明のさらに別の目的は、スイッチ内にプロトコルを解釈する手段をつけることによりスイッチで上位装置のアクセスの許可を判定する技術を提供することにある。本発明のさらに別の目的は、スイッチ内にプロトコルを解釈する手段をつけることによりスイッチで上位装置のコマンドの内容を理解する技術を提供することにある。本発明のさらに別の目的は、スイッチ内にパケット内容を変更する手段をつけることにより、パケットの送信先を変更する技術を提供することにある。本発明のさらに別の目的は、スイッチ内にキャッシュ処理を行うパス、プロトコ

ル変換を行うパスを別々に持つことにより、キャッシュ使用、不使用の切り替えを行う技術を提供することにある。

【0007】

【課題を解決するための手段】上記目的を達成するため、本発明は、複数の上位装置と、磁気ディスク、テープライブラリ等の記憶装置にパスで接続され、前記上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチであり、前記記憶装置の構成情報を記憶する手段と、該構成情報の管理及びアクセス権の管理を行う構成情報管理・アクセス権管理手段と、キャッシュメモリと、該キャッシュメモリを管理するキャッシュ管理手段と、前記パスの切り替えを行うパス切り替え手段を備えるようにしている。また、該スイッチで受信、送信を行うフレームを外部の接続装置にあったプロトコルに変換するプロトコル変換手段をさらに備えるようにしている。また、該スイッチに該スイッチ内のキャッシュメモリ、またはプロトコル変換手段を使用しないパスも接続するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレームの中の上位プロトコル情報を取り出し、次の処理を決定するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレームの中の上位プロトコル内にあるペイロードの情報を取り出し、該情報に基づき前記キャッシュメモリ内にデータがあるかどうかを判定するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレームのヘッダ内にある出力先アドレスを書き換えて、書き換えられた出力先アドレスを基に出力側ポートを決定するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレーム内のアドレスおよび上位プロトコル内のペイロード内にある上位レイヤシーケンスが、既に記憶してあるアドレスおよび上位レイヤシーケンスの範囲内であるか否かを判定し、あると判定したとき、受信したデータを前記キャッシュ管理手段によりキャッシュメモリにも格納するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレーム内の一部の情報がすでに記憶してある情報の一部に一致するか判定し、一致すると判定したとき、受信したデータを前記キャッシュ管理手段によりキャッシュメモリにも格納するようにしている。また、該スイッチ内のキャッシュメモリのデータを外部から与えられたタイミングで前記キャッシュ管理手段により読み出し、該読み出したデータをパス切り替え手段で選択したパスを介して任意の磁気ディスクやテープライブラリに書き込むようにしている。また、パスを介して転送された磁気ディスクやテープライブラリの任意のデータを外部から与えられたタイミングで該スイッチ内のキャッシュメモリに前記キャッシュ管理手段により格納するようにしている。また、前記構成

情報管理・アクセス権管理手段は、該スイッチに接続されている複数の上位装置、磁気ディスク、テープライブラリからスイッチに内蔵されているキャッシュメモリへの複数のアクセスを並列に処理するようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチ内部のキャッシュメモリに格納するデータを暗号化し、許可されていないアクセスに対してデータの解読を防ぐようにしている。また、前記記憶装置の構成情報を記憶する手段に該スイッチ内部のキャッシュメモリに格納するデータに対して論理Vol.毎に暗号化鍵を記憶し、前記構成情報管理・アクセス権管理手段は、論理Vol.毎に暗号化を行うようにしている。また、前記構成情報管理・アクセス権管理手段は、該スイッチで受信したフレームの中の上位プロトコル内にあるペイロードの情報を取り出し、該情報に基づき前記キャッシュメモリ内にデータがあるかどうかを判定し、キャッシュメモリ内にデータがあると判定した場合、前記キャッシュ管理手段によりキャッシュメモリよりデータを読み出し、読み出したデータを該データを請求した上位装置に送信するようにしている。また、前記構成情報管理・アクセス権管理手段は、前記キャッシュメモリ内にデータがあるかどうかの判定で、キャッシュメモリ内にデータがないと判定した場合、該データを請求したフレームのプロトコル送信IDを記憶するようにしている。また、前記構成情報管理・アクセス権管理手段は、前記データを請求したフレームのプロトコル送信IDと一致するフレームが磁気ディスクまたはテープライブラリから受信された場合、該フレーム内のデータを前記キャッシュ管理手段によりキャッシュメモリに書きこんだ後上位装置に転送するようにしている。

【0008】また、複数の上位装置と、磁気ディスク、テープライブラリ等の複数の記憶装置と、該上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチを備える計算機システムであり、前記キャッシュ内蔵型スイッチは、前記記憶装置の構成情報を記憶する手段と、該構成情報の管理及びアクセス権の管理を行う構成情報管理・アクセス権管理手段と、キャッシュメモリと、該キャッシュメモリを管理するキャッシュ管理手段と、前記パスの切り替えを行うパス切り替え手段を備えるようにしている。さらに、前記キャッシュ内蔵型スイッチで受信するフレームを該キャッシュ内蔵型スイッチにおけるプロトコルに変換し、該キャッシュ内蔵型スイッチ送信から送信するフレームを外部の接続装置にあったプロトコルに変換するプロトコル変換手段をさらに備えるようにしている。

【0009】また、複数の上位装置と、磁気ディスク、テープライブラリ等の記憶装置にパスで接続され、前記上位装置と記憶装置間の接続の切り替えを行うキャッシュ内蔵型スイッチのスイッチ制御方法であり、受信したフレームから発信元ID、発信先IDを取得し、前記キ

キャッシュ内蔵型スイッチ内に設けられた構成情報テーブルを参照して発信元IDが発信先IDに対するアクセス権を有する場合、該フレーム内の情報が論理Vol.アクセスを指定しているか否かの判定をし、指定していない場合はパス切り替え処理を行い、該フレームをフレームで指定された記憶装置に送信し、指定している場合には、前記フレームから論理Vol.とアドレスを取り出し、前記構成情報テーブルを参照して該論理Vol.とアドレスを書き換える実送信IDとアドレスを求め、書き換えを行い、前記キャッシュ内蔵型スイッチ内に設けられたキャッシュテーブルを前記論理Vol.とアドレスにより参照してキャッシュヒットの場合、データをアクセスし、該データを発信元に送信するためパス切り替え処理を行い、キャッシュミスの場合、パス切り替え処理を行い、前記書き換えたフレームを実送信先の記憶装置に送信し、記憶装置から返送されたアクセスされたデータをキャッシュメモリに格納し、該データを発信元に送信するためパス切り替え処理を行うようにしている。さらに、前記受信したフレームのプロトコルが前記キャッシュ内蔵型スイッチにおけるプロトコルと異なる場合には、該受信したフレームのプロトコルを該キャッシュ内蔵型スイッチにおけるプロトコルに変換し、前記キャッシュ内蔵型スイッチから送信するフレームが送信先の装置のプロトコルと異なる場合には、該キャッシュ内蔵型スイッチから送信するフレームのプロトコルを該送信先の装置のプロトコルに変換するようにしている。

【0010】

【発明の実施の形態】以下、本発明に係る実施例を図面に示しさらに詳細に説明する。図1は本発明のキャッシュ内蔵型スイッチの実施例における全体構成図である。10、15、20、25は上位装置であり、磁気ディスク125～170、テープライブラリ120等の記憶装置に対して読み出し、書きこみ等の命令を発行する。30、35、40、45は上位装置10～15とスイッチ60を接続する外部バスであり、現在オープン用としてファイバーチャネル(FC)、メインフレーム用としてエスコンチャネル(ESCON)がすでに製品化されている。50はLANに代表される他のネットワークであり、こことスイッチ60間でフレームを送受信できる。また97、100、105、110、115は磁気ディスク125～170、テープライブラリ120等の記憶装置とスイッチ60を接続する外部バスであり、現在SCSI、FC等が製品化されている。55はスイッチ60を制御するサーバプロセス(SVP)でありスイッチ内の初期値の設定、動作の確認、障害時の処理を行う。

【0011】60は上位装置10～25と磁気ディスク125～170、テープライブラリ120等の記憶装置を接続するスイッチであり、装置間の接続の切り替え、コマンド、データの転送、さらにキャッシュメモリの管理、磁気ディスク等の論理ボリューム(論理Vol.)管理を

制御する。65、70はプロトコル変換部であり、スイッチで接続された装置間のプロトコルの変換等を行い、パケット内の情報の読み出し、更新を行う。75はバス切り替え部であり、あらかじめ設定された論理に従って、バス30～45、97～115間の接続切り替えを行う。80は構成情報管理、アクセス権管理部であり、スイッチ60に接続された磁気ディスク125～170、テープライブラリ120等の記憶装置の管理、記憶装置間にまたがる論理Vol. 185～190等の構成情報の管理、キャッシュメモリ95に格納されているデータの管理を行う。またパケットによってはその内容の変更を行う。さらに上位装置10、15、20、25のアクセス権も管理している。

【0012】85は構成情報テーブルであり、スイッチ60に接続された記憶装置の管理情報、キャッシュメモリ95の管理情報を記憶している。90はキャッシュ情報管理部であり、キャッシュヒットの場合はデータを読み出し、キャッシュミスの場合はデータの磁気ディスク125～170、テープライブラリ120等の記憶装置への書き戻しと該当するデータの読み出しを行う。キャッシュ制御のアルゴリズムはLRU(Least Recently Used)等を用いているが、この技術はすでに公知なので特に説明は行わない。95はキャッシュメモリであり、磁気ディスク125～170、テープライブラリ120等の記憶装置データの中で使用頻度が高いものを一時的に格納する。

【0013】上記スイッチ60内のデータの流れは上位装置10～25と磁気ディスク125～170、テープライブラリ120の組み合わせによって変わる。フレーム400(図4)がキャッシュメモリ95もプロトコルの変換も必要ない場合はバス切り替え部75のみ通る

(①)。フレーム400がキャッシュメモリ95を使用する場合はバス切り替え部75、構成情報管理、アクセス権管理部80を通る。(②)

フレーム400がプロトコル変換、キャッシュメモリ95を使用する場合は、バス切り替え部75、構成情報管理、アクセス権管理部80、およびプロトコル変換部65または70を通る。(③)

このようにして各々のデータの性質によって異なるバスを通ることで、応答時間の最適化を図ることが可能となる。

【0014】120はテープライブラリであり、比較的使用頻度の低いデータを保存している。125、130、135、140、155、160、165、170は磁気ディスクであり、データを格納している。180、185、190は論理Vol. であり、上位装置10～25の一部のトランザクションアドレスはこの論理ボリュームを指している。この論理Vol. 180、185、190はスイッチ内の構成情報テーブル85が管理している。

【0015】図3に本発明の実施例におけるスイッチ内部の構成図を示す。300はクロスバススイッチであり、バス切り替え部で最終的に決定した組み合わせに従ってバスを接続する。305は全体管理プロセッサであり、プロトコル変換部70、構成情報管理、アクセス権管理部80、バス切り替え部75、構成情報テーブル85をプログラムに従って制御する。310、315、320は入出力部であり、入力または出力されるデータのバスが接続されている。311、316、321は出力側バスでありここから外部にデータが送信される。312、317、322は入力側バスであり外部からデータがここに受信される。図3の例では入力側バス312に入るフレームは直接バス切り替え部75に入りクロスバススイッチで切り替えられるような接続にされている。また、入力側バス315はプロトコル変換部70を通ることなく構成情報管理、アクセス権管理部80に入る接続にされている。出力311、316はクロスバススイッチ300から直接フレーム400を受信する接続にされている。

【0016】図4に本発明の実施例におけるフレーム構造例を示す。この構造例はファイバチャネルの標準規格のフォーマットを参考にしたものである。400はフレームの例であり、幾つかの部分から成り立つデータである。405はStart Of Frame識別子であり、フレームの先頭をスイッチ60の入出力部310～320に伝える。410はフレームヘッダであり、フレームのアドレス等を記述している。詳細は後で述べる。415は、ペイロードであり上位プロトコルのデータ、コマンド等が格納されている。詳細は後で述べる。420はCRC(Cyclic Redundancy Check)であり、フレーム400内のデータの誤りを発見し訂正を行う。425はEnd Of Frame識別子であり、フレームの終わりをスイッチ60の入出力部310～320に伝える。

【0017】次にフレームヘッダ410の構造例を示す。430は送信先ポートIDでありフレーム400が次に進みたい装置のポートIDを示す。435は送信元ポートIDでありフレームをスイッチ60に送信してきた装置のポートIDを示す。440は上位レイヤプロトコルであり、送信元と送信先がデータのやり取りを行う際に使うプロトコルの種類を示しており、たとえばSCSI等があげられる。445は上位レイヤシーケンスIDであり、上位プロトコルに従った通信におけるシーケンスの順番を示す。450はシーケンス内フレーム番号であり、シーケンス内にある複数のフレームの順番を示す。455は上位レイヤプロトコル送信先IDであり、最終的にフレーム400が到着する装置のポートIDを示す。460は上位レイヤプロトコル送信元IDであり、フレームを最初に送信した装置のポートIDを示す。

【0018】次にペイロード415の構造例を示す。ここで例としてリードを行うコマンドを挙げる。475は

オペレーションコードであり、上位プロトコルで使われるコマンドの種類を示しており、この例ではリードコマンドとなる。480は論理Vol. 番号であり、例えば磁気ディスク125～170に設定された 論理Vol. 180、185、190を示している。485はデータ等が存在する場所のアドレスであり、論理Vol. 180、185、190を用いた場合は論理Vol. 番号480内のアドレスを示し、その他の場合は装置のIDと装置内のアドレスを示している。490は転送データ長であり、リードする範囲を示す。495はコントロールバイトであり、フレーム400またはペイロード415が扱える容量に合わせるために付け加える空白なデータである。

【0019】図5に本発明の実施例における構成情報テーブル構造およびキャッシュテーブルを示す。500は構成情報テーブル#1であり各コマンドの発行元のアクセス権等を管理している。505は上位レイヤプロトコル発信先IDであり、スイッチ60にきたフレーム400が持つ、上位レイヤプロトコル送信先ID455に対応している。510はアクセス可能送信元IDであり、上位レイヤプロトコル送信先ID505が許可している送信元ID 20を示しており、ここに書かれた送信元IDのみが上位レイヤプロトコル発信先ID505にアクセスすることが許可されている。515は送信先ID/ポインタを示しており、実際にアクセスするための送信先のIDもしくは、該当する送信先が含まれる論理Vol. 180、185、190またはキャッシュメモリ95を管理している構成情報テーブル#2 520へのポインタを示している。520は構成情報テーブル#2であり、論理Vol. 180、185、190に含まれるデータをアクセスする際にそのアドレス等を管理する。525はアクセス先論理 30 Vol. 番号であり、指定した論理Vol. 番号480が示してある。530はアドレス範囲であり、その論理Vol. 180、185、190で扱うデータのアドレスの範囲を示している。同じ論理Vol. 180、185、190でも物理的に異なる磁気ディスク125～170にデータがわかれている場合にはこのデータで判別する。535はアクセス先ロックフラグであり、すでにスイッチ60内に入ったコマンドがその論理Vol. をアクセスしていた場合には、このフラグを立てて後から来たコマンドのアクセスを待たせる。540は実送信先IDであり、実際に 40 アクセスするデータが存在する記憶装置のIDを示している。545はキャッシュ使用フラグであり、もしアクセスするデータがキャッシュを使っているならばこのフラグを立てることにより、キャッシュテーブル550をアクセスできるようにする。

【0020】550はキャッシュテーブルであり、キャッシュメモリ95にアクセスするデータが存在しているかを示している。555はアクセス先論理Vol. 番号でありアクセス先論理Vol. 番号525と同じものである。560はアクセス先アドレスであり、アクセスするデータ 50

の磁気ディスクにおけるアドレスを示す。565はキャッシュメモリアドレスであり、キャッシュメモリ95のデータのアドレスを示す。570はキャッシュ無効フラグであり、アクセスするデータが書きかえられたりして、古い場合にはこのフラグが立つ。575はキャッシュロックフラグであり他のアクセスによってこのデータのキャッシュが使用中である場合、このフラグを立てる。なお構成情報テーブル#1 500、構成情報テーブル#2 520はスイッチ60内の構成情報テーブル85内に格納される。キャッシュテーブル550はキャッシュ情報管理部90または構成情報テーブル85内に格納される。

【0021】図6に本発明の実施例におけるリードキャッシュヒット時のデータの流れを示す。上位装置10～25の内の、例えば上位装置20よりリードコマンドがスイッチ60に発行される(①)。次にプロトコル変換部65でフレーム400内のペイロード415を解析し 10 わかった結果を構成情報管理、アクセス権管理部80に渡し、アクセス権等の確認を行う。また論理Vol. 番号480、アドレス485と構成情報テーブル85を比較する(②)。比較した結果、キャッシュヒットと判定された場合、キャッシュメモリ95よりデータを読み出す

(③)。次に構成情報管理、アクセス権管理部80で上位装置20にキャッシュヒットしたデータを返すようなフレームを作成し、次に、パス切り替え部75にてフレームのアドレスに従って、上位装置20のポートにフレームを送信できるようにクロスバススイッチ300を切り替える(④)。その後、上位装置20にデータが送信される(⑤)。

【0022】図7に本発明の実施例におけるリードキャッシュミス時のデータの流れを示す。上位装置10～25の内の、例えば上位装置20よりリードコマンドがスイッチ60に発行される(①)。次にプロトコル変換部65でフレーム400内のペイロード415を解析した結果を構成情報管理、アクセス権管理部80に渡し、アクセス権等の確認を行う。また論理Vol. 番号480、アドレス485と構成情報テーブル85を比較する

(②)。比較した結果、キャッシュミスと判定された場合、構成情報テーブル#2 520からわかる実送信先IDをフレーム400のアドレス485に書きこみ、それに従って、パス切り替え部で必要なポートにフレームを送信できるようにクロスバススイッチ300を切り替える(③)。次にデータのある磁気ディスク130よりデータを読み出し、データのコピーはキャッシュメモリ95に格納、キャッシュテーブル550を変更した後

(④)、上位装置20のポートにフレームを送信できるようにクロスバススイッチ300を切り替える(⑤)。その後、上位装置20にデータが送信される(⑥)。

【0023】図8に本発明の実施例におけるプロトコル変換処理のフローチャートを示す。ここではプロトコル

変換部65~70と外部バス30、35、40、45間の処理のフローである。ステップ600のプロトコル変換処理開始から処理が始まる。ステップ605では行う処理が出力側バス311、316、321か入力側バス312、317、322かの判定を行う。出力側バス311、316、321の処理であれば、ステップ630に進み、バス切り替え部75よりデータを受信する。ステップ635では出力するためにフレームヘッダ410情報を作成する。ステップ640ではプロトコル変換部65~70で送信先のフォーマットに変換を行う。ステップ645では出力側バス311、316、321へデータを転送してステップ605に戻る。入力側バス312、317、322の処理であれば、ステップ610に進み、入力側バス312、317、322よりパケットを受信する。ステップ615では発信元にパケットを受領したことを通知する。ステップ620ではデータの変換を行い、フレームヘッダ410やペイロード415をパケットより抽出し内容の解析を行う。ステップ625では構成情報管理、アクセス権管理部80に解析結果とペイロードを転送し、ステップ605に戻る。

【0024】図9に本発明の実施例における構成情報管理、アクセス権管理部処理のフローチャートを示す。ここでは、構成情報管理、アクセス権管理部80が受信したフレーム400の情報から送信先のポートの決定またはキャッシュヒット処理への移行を行う。ステップ700の構成情報管理、アクセス権管理部処理開始より処理が始まる。ステップ705では上位レイヤプロトコル送信先ID455、上位レイヤプロトコル送信元ID460をフレームヘッダ410より取り出す。ステップ710では構成情報テーブル#1500の上位レイヤプロトコル送信先ID505をアクセスし、発信可能な発信元IDを探す。ステップ715では上位レイヤプロトコル送信元ID460が上位レイヤプロトコル送信先ID455をアクセス可能であるかを判定する。もしアクセス不可であれば、アクセスエラーとなる。ここではエラー処理が発生するが、本発明には本質的でないのでこれ以上の説明は省略する。もし可能であればステップ725に進み、上位レイヤプロトコルが論理Vol.180、185、190のアクセスであるかを判定を行う。論理Vol.180、185、190のアクセスで無ければ（例えば、テープライブラリあるいは論理Volに入っていない磁気ディスクなどへのアクセスである場合）、ステップ730に進みバス切り替え処理を行う。論理Vol.180、185、190のアクセスであればステップ735に進み、ペイロード415内の論理Vol.番号480とアドレス485を取り出す。ステップ740では構成情報テーブル#2520から実送信先ID540を決定し、また送信先の実アドレスを実送信先ID540とアドレス485から計算し、ペイロード内の論理Vol.番号を実送信先IDに、アドレス485を送信先の実アドレスにそれぞれ置

き換える。ステップ745ではアクセス先ロックフラグ535を調べ、アクセス先がすでに使用中かどうか調べる。もし使用中であれば、そのまま待つ。使用中でなければ、ステップ746でアクセス先ロックフラグを1にしてキャッシュヒット処理750へ進む。

【0025】図10に本発明の実施例におけるキャッシュヒット処理のフローチャートを示す。ここではキャッシュメモリ95にアクセスするデータが存在するかを判定する。ステップ750のキャッシュヒット処理より処理が始まる。ステップ755では構成情報テーブル#2520よりアクセスするデータのキャッシュ使用フラグ545を取り出す。ステップ760ではキャッシュ使用フラグ545を見てキャッシュメモリ95を使用しているかどうかを判定する。もしキャッシュメモリ95を使用していなければ、ステップ800のバス切り替え処理に進む。もしキャッシュメモリ95を使用していればステップ765に進み、該当するデータについてキャッシュテーブル550内のアクセス先論理Vol.番号555およびアクセス先アドレス560からキャッシュ無効フラグ570、キャッシュロックフラグ575を取り出す。ステップ770ではキャッシュテーブル550に該当するデータがあり、キャッシュ無効フラグが1でないキャッシュヒットの状態であるかを判定する。もしキャッシュヒットであればステップ775に進む。ステップ775ではキャッシュロックフラグ575を調べ、アクセス先がすでに使用中かどうか調べる。もし使用中であれば、そのまま待つ。ステップ780ではキャッシュロックフラグ575を設定し、キャッシュメモリ95のデータをアクセスする。ステップ781ではアクセス先ロックフラグ535とキャッシュロックフラグ575を解除する。ステップ785ではヘッダ内の上位レイヤプロトコル送信先ID455を自分のポート（入力ポート）、上位レイヤプロトコル送信元IDをスイッチに変更し、データを、命令を送信した上位装置20へ返す。そしてステップ800のバス切り替え処理に進む。もしステップ770でキャッシュヒットしていなければ、ステップ790に進み、キャッシュテーブル550に新たにエントリを作り、このエントリのキャッシュ無効フラグ570を設定する。ステップ795では上位レイヤプロトコル送信先ID455を記憶する。ステップ800ではバス切り替え処理を行う。ステップ900ではキャッシュミス処理を行う。

【0026】図11は、本発明の実施例におけるバス切り替え処理のフローチャートを示す。ステップ800のバス切り替え処理から始まる。ステップ805では一定時間バス切り替え要求があるかを監視する。ステップ810では、各出力側バス311、316、321毎に使用入力側バス312、317、322を決定する。決定方式はラウンドロビン方式等があるが、公知の技術なのでここでは説明を省略する。ステップ815で発信元は

スイッチ60であるかを判定する。スイッチ60であればステップ820に進み、パスを切り替え、他のトランザクションが使えない様にロックする。ステップ825ではリードまたはライト処理に必要なプロトコルを実行する。ステップ830ではトランザクションの転送、およびロックの解除を行う。ステップ840では次のスイッチ60切り替え処理の優先順位を再計算する。またアクセス先のロックを解除する。ステップ845では、パス切り替え処理を終了する。ステップ815で発信元はスイッチ60でなければステップ835に進み、パスを切り替え、トランザクションの転送を行う。次にステップ840に進む。

【0027】図12は、本発明の実施例におけるキャッシュミス処理のフローチャートを示す。ステップ900からキャッシュミス処理が始まる。ステップ905では、図10のステップ795で記憶した上位レイヤプロトコルID455を持つトランザクションであるかどうかの判定を行う。もし上位レイヤプロトコルID455を持つトランザクションでなければ、来るまで待つ。もし上位レイヤプロトコルID455を持つトランザクションならば、ステップ910に進み、キャッシュメモリ95にデータを格納する。ステップ915では構成情報テーブル#2520、キャッシュテーブル550に管理情報を登録し、キャッシュメモリ95を有効にする。ステップ916で記憶した上位レイヤプロトコルID455を消去する。ステップ920でパス切り替え処理800(図11)に進む。以上の説明ではキャッシュメモリのライトはデータがキャッシュに存在しない限り行わない制御方式(ライトスルー方式)で説明したが、データのライトは必ずキャッシュにする方式(ライトアフター方式)等の他の制御方式を使っても本発明の目的と矛盾するところはなく、本発明に含まれる。また上位装置10~25のタイミングを待たずにスイッチ60内の制御プログラムによって、キャッシュメモリ95内のデータを強制的に磁気ディスク125~170やテープライブラリ120に格納すること、また磁気ディスク125~170からテープライブラリ120にデータを移動する際にキャッシュメモリ95をバッファとして使うことも自明である。

【0028】本発明を用いた場合の効果を図13に示す。図13(a)に示したようにスイッチにキャッシュが内蔵されていない場合はサーバ950とスイッチ955間のパケットのやり取りに加え、スイッチ955と磁気ディスクまたはキャッシュメモリ960の間で同様なパケットのやり取りを行う必要がある。そのため応答時間に余計なオーバーヘッドがかかってしまう。しかし図13(b)に示したように、スイッチ970内にキャッシュメモリが存在し、ヒットした場合には図13(a)に示したようなオーバーヘッドは無いのでその分応答時間が短縮される。

【0029】図14に本発明の実施例における暗号化部の処理の流れを示す。キャッシュには各論理論理Vol.のデータが混在することになるので、論理Vol.にアクセス権のない上位装置10~25、SVP55等が誤ってデータを読み出す可能性がある。そのためにキャッシュ内のデータは暗号化されていることが望ましい。本発明では、公開鍵暗号の技術を使って上位装置10には復号部980および復号鍵981(秘密鍵)、スイッチ60には暗号部985および暗号化鍵984(公開鍵)を持つことによりデータを暗号化し、信頼性を高める。またコマンド、データ内で暗号化の実行を表示するビットは図15(a)、図15(b)で示す。なお、公開鍵暗号の技術はすでに公知の技術なので詳細な説明は省略する。上位装置10はアクセス可能な論理ボリューム毎に復号化鍵981を設定する。また上位装置10はデータ(暗号化)990を元に戻すための復号化部980を持つ。スイッチ60は構成情報管理、アクセス権管理部80内に暗号化部985を持つ。構成情報テーブル85内には、暗号化テーブル982を持ち、内容は論理Vol.名983と暗号化鍵984から成り立つ。暗号化テーブル982内の暗号化鍵984は初期化の段階で各論理Vol.へのアクセスが許可されている上位装置より受け取る。

【0030】上位装置10から発行されたコマンド+暗号化415(①)は暗号化ビット495がオンならば、構成情報管理、アクセス権管理部80内の暗号化部にデータの暗号化を指示する(②)。暗号化部はコマンドがリード命令の場合はデータをキャッシュメモリ95、磁気ディスク125~170、テープライブラリ120より読み出し暗号化した後データ内の暗号化ビット499をオンにしてキャッシュメモリ95に格納する(③)。ただし読み出したデータの暗号化フラグ499がすでにオンの場合は暗号化を行わない。データ(暗号化)988はバス切り替え部75にも渡される(④)。その後クロスバススイッチ300、プロトコル変換部70、入出力部320を通して上位装置10に転送される(⑤)。上位装置では復号鍵981を使って復号化部980で元のデータ(暗号化なし)991に変換される(⑥)。コマンドがライトの場合はデータを暗号化部988で暗号化し(②)、暗号化フラグをオンにした後キャッシュメモリ95に格納する(③)。

【0031】図15(a)に本発明の実施例における暗号化用ペイロード例を示す。ペイロード415のコントロールバイト495のビット(リザーブ領域)を用いて暗号化ビット496とする。1の場合は暗号化ビット496がオンとなり、データを暗号化するように暗号化部988に指示する。図15(b)に本発明の実施例におけるデータ+暗号化例を示す。データ+暗号化987は先頭のビットが暗号化ビット499であり、1の場合暗号化がオンとなり以下のデータは暗号化されていることになる。なお図15(a)、図15(b)で説明したペ

イロード例、データ+暗号化例以外にも暗号化を行うプロトコルが検討されている。しかしそれらも本発明の範囲内であることは自明である。

【0032】以上図1から図15までに示した実施例によれば、上位装置10～25および磁気ディスク125から170、テープライブラリ120等を接続するスイッチ60において、スイッチ60内部で構成情報テーブル85およびキャッシュメモリ95を制御することにより、リード、ライトがヒットした場合はいったん外部の記憶装置までトランザクションを送る必要がなくなり、各処理待ち時間が短縮し、処理トランザクション数の増加を図ることが可能になる。さらにこの実施例にあるスイッチ60を多数接続することにより、上位装置10～25および磁気ディスク125から170、テープライブラリ120等の増加に応じて、キャッシュメモリ95の容量を増加することが可能となるので簡単にキャッシュヒット率を維持していくことが可能となる。またスイッチ60を階層的に接続することにより、必要なデータのみ上位装置の近辺のスイッチに置けるので、ネットワーク全体のトラフィック量を削減し性能が向上する可能性がある。さらにこの実施例にあるキャッシュメモリ95をバッファにして磁気ディスク125～170からテープライブラリ120にデータを移動することで上位装置に負荷をかけることなくデータの移動およびバックアップを行うことが可能となる。さらにこの実施例のスイッチ60を用いることにより、外部接続した装置のプロトコルに合わせてフレーム400を変換していけるので、外部に接続可能な装置の自由度を増やすことが可能となる。またキャッシュ内のデータは暗号化できるので、データの信頼性を向上可能となる。

【0033】

【発明の効果】スイッチに接続され、スイッチ内のキャッシュメモリの管理している磁気ディスクへのアクセスの中で、キャッシュ内にデータが存在する場合、磁気ディスクまたはスイッチに接続されている外部メモリまでアクセスすることなくデータを送信できるので、トランザクションの応答時間が短縮され、記憶装置の高性能化が図れる。

【図面の簡単な説明】

【図1】本発明の実施例における全体構成を示す図である。

【図2】従来のスイッチ接続型記憶装置の構成を示す図である。

【図3】本発明の実施例におけるスイッチ内部の構成を示す図である。

【図4】本発明の実施例におけるフレーム構造例を示す

図である。

【図5】本発明の実施例における構成情報テーブル構造およびキャッシュテーブルを示す図である。

【図6】本発明の実施例におけるキャッシュヒット時のデータの流れを示す図である。

【図7】本発明の実施例におけるキャッシュミス時のデータの流れを示す図である。

【図8】本発明の実施例におけるプロトコル変換処理のフローチャートを示す図である。

10 【図9】本発明の実施例における構成情報管理、アクセス権管理部処理のフローチャートを示す図である。

【図10】本発明の実施例におけるキャッシュ処理のフローチャートを示す図である。

【図11】本発明の実施例におけるパス切り替え処理のフローチャートを示す図である。

【図12】本発明の実施例におけるキャッシュミス処理のフローチャートを示す図である。

【図13】本発明の実施例におけるスイッチと上位装置、磁気ディスク間のプロトコルを示す図である。

20 【図14】本発明の実施例における暗号化部の処理の流れを示す図である。

【図15】本発明の実施例における暗号化用ペイロードとデータの例を示す図である。

【符号の説明】

10、15、20、25 上位装置

50 他のネットワーク

55 SVP

60 スイッチ

65、70 プロトコル変換部

30 75 パス切り替え部

80 構成情報管理、アクセス権管理部

85 構成情報テーブル

90 キャッシュ情報管理部

95 キャッシュメモリ

120 テープライブラリ

125、130、135、140、155、160、1

65、170 磁気ディスク

180 論理Vol. 1

185 論理Vol. 2

190 論理Vol. 3

300 クロスバススイッチ

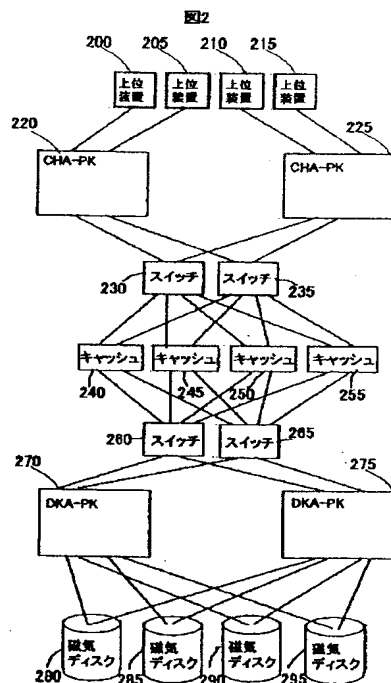
305 全体管理プロセッサ

310、315、320 入出力部

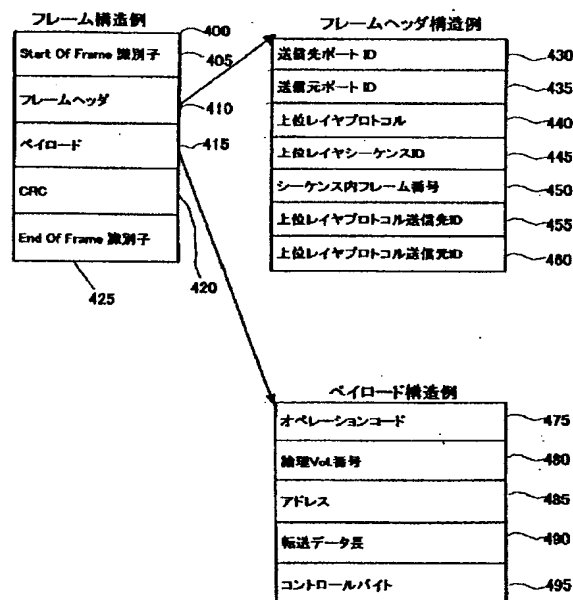
311、316、321 出力

312、317、322 入力

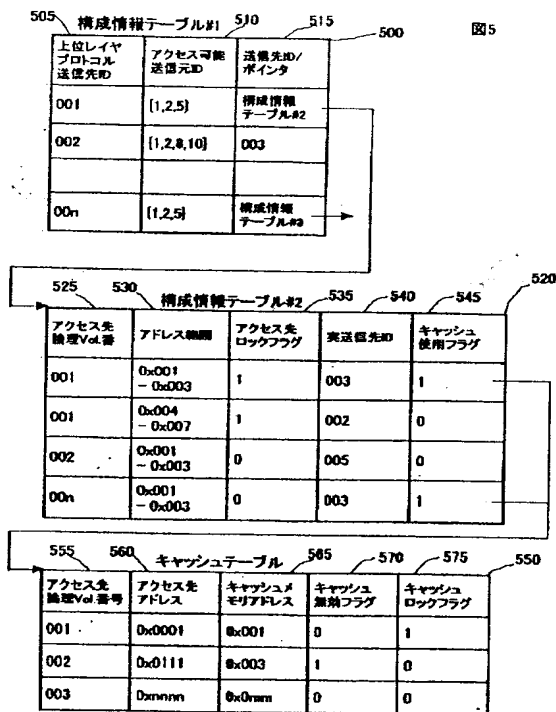
【图2】



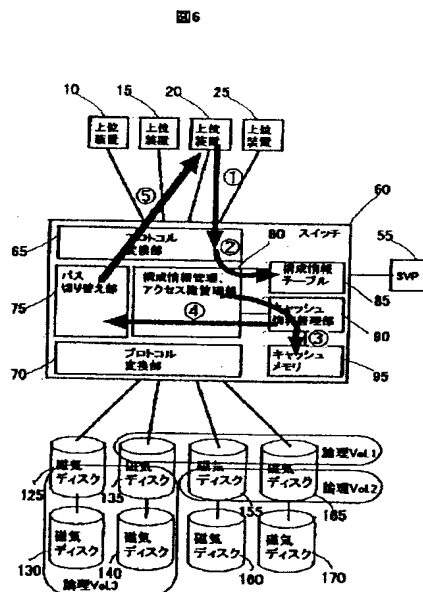
【図 4】



【図5】

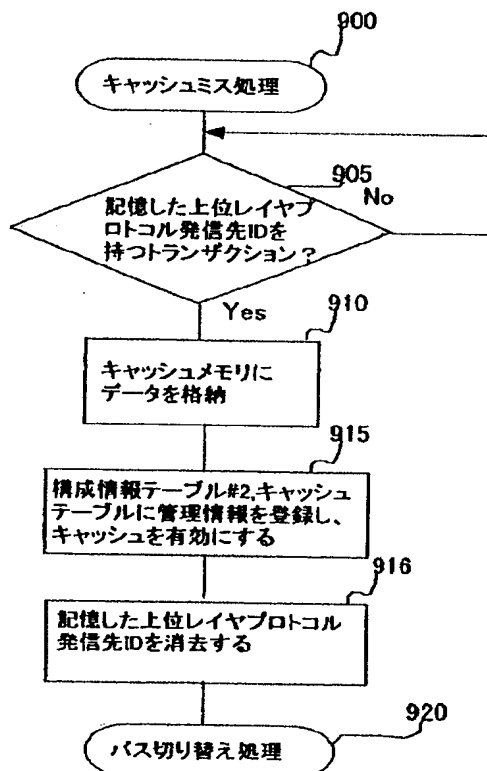
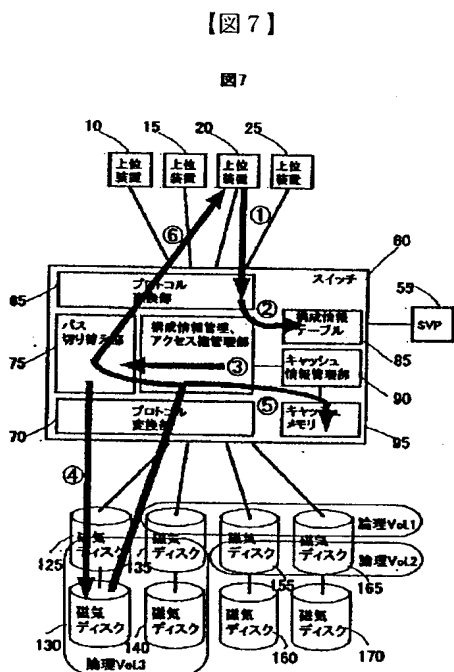


【図6】

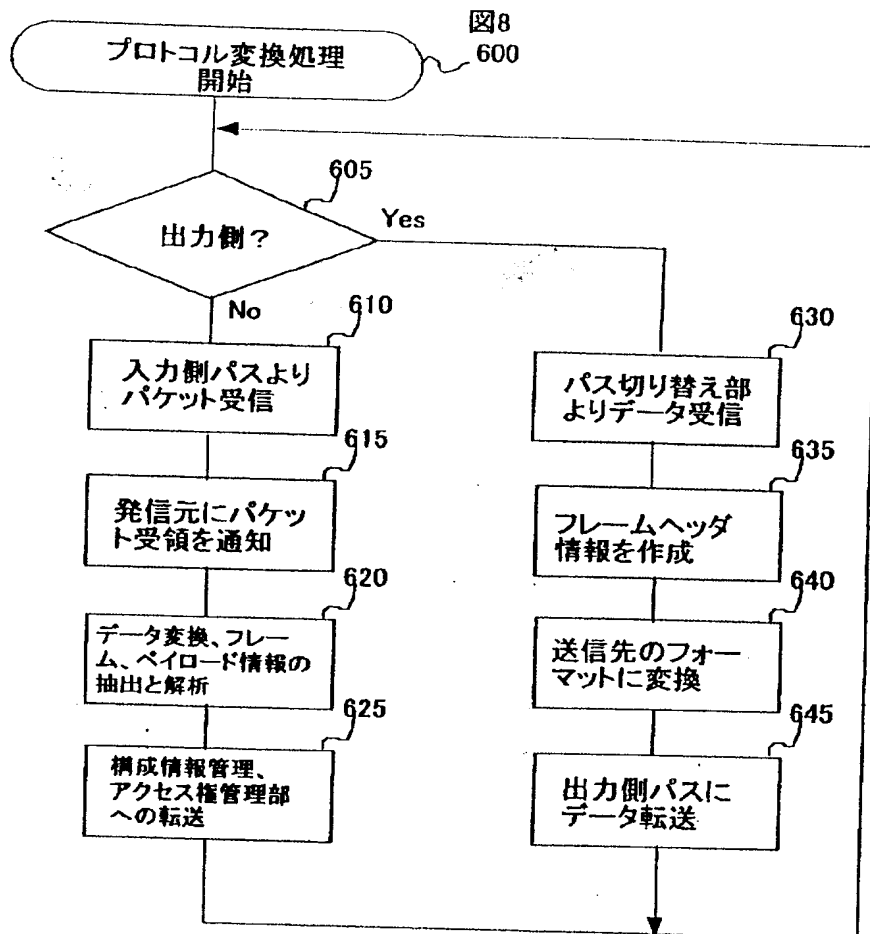


【図12】

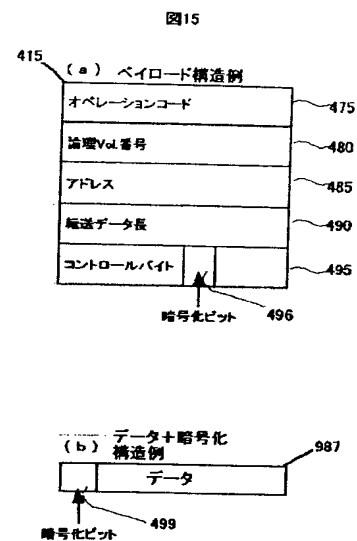
図12



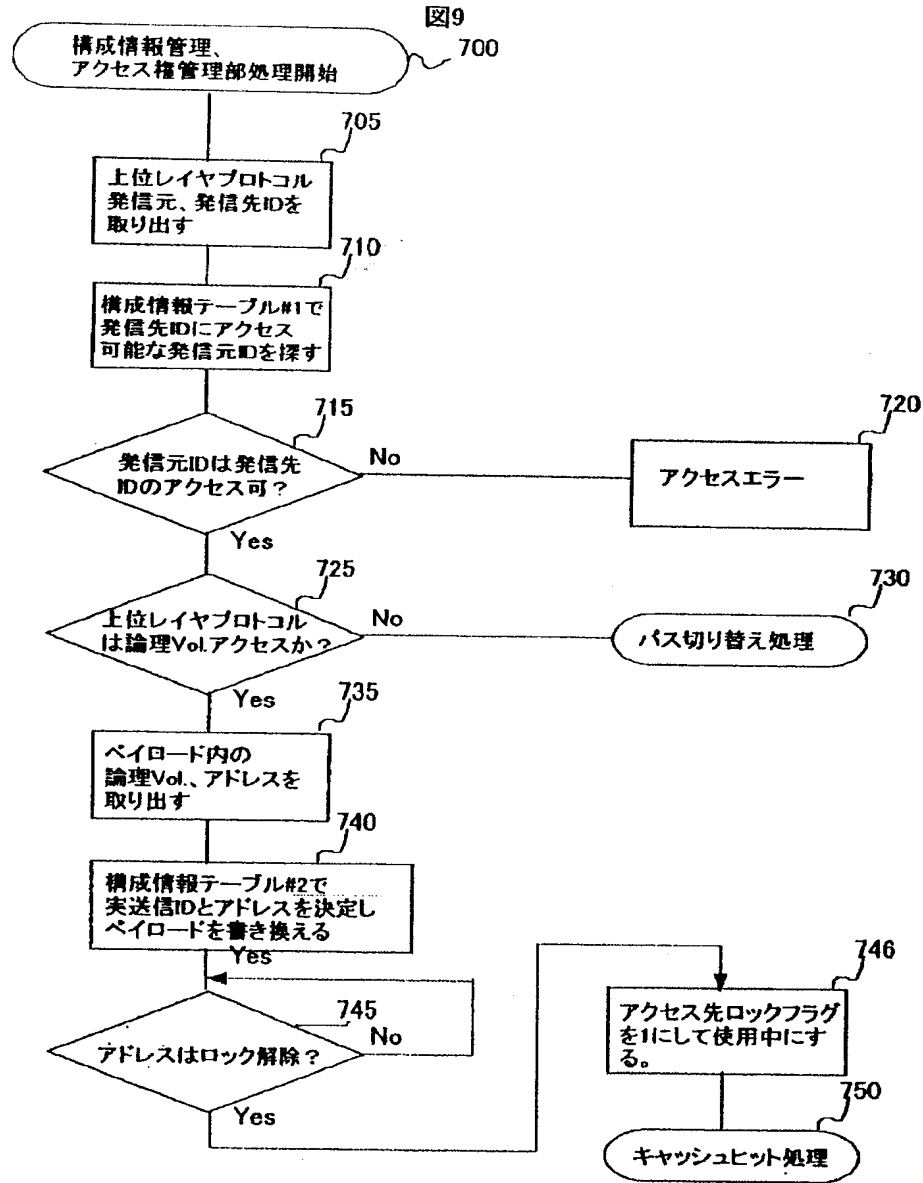
【図8】



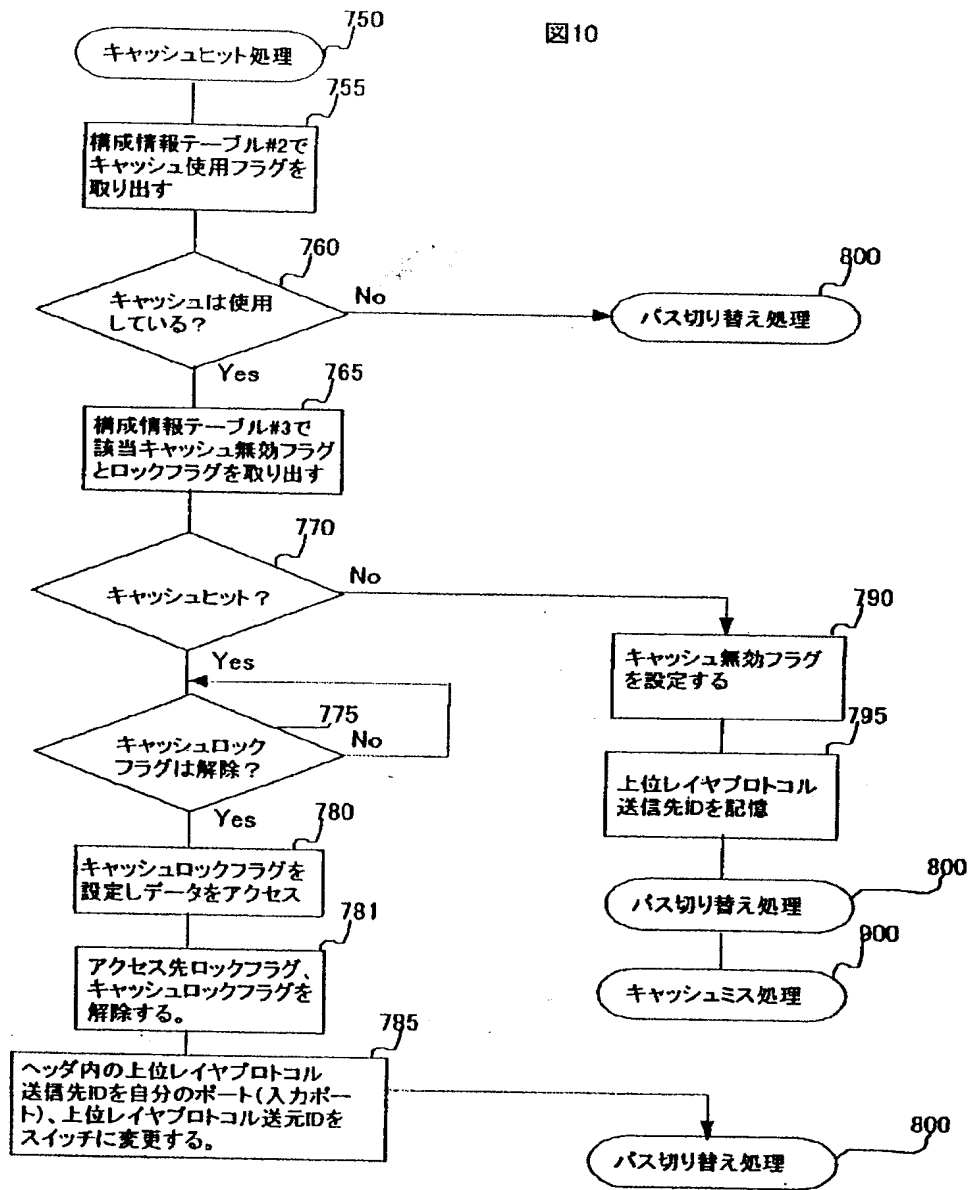
【図15】



【図9】

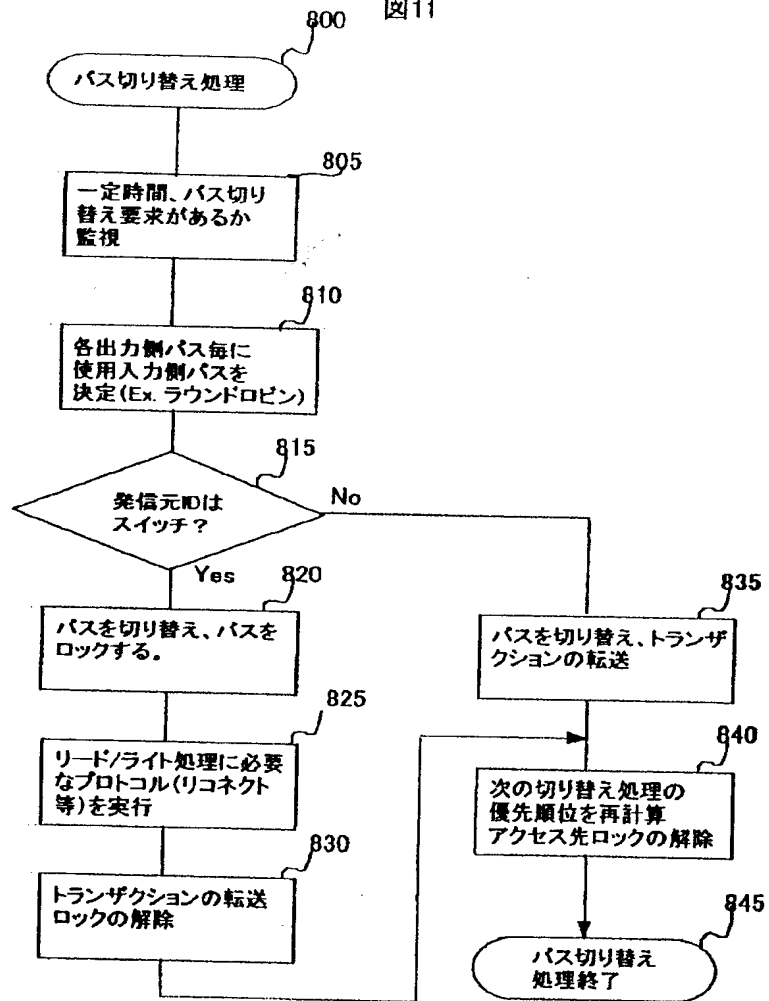


【図10】

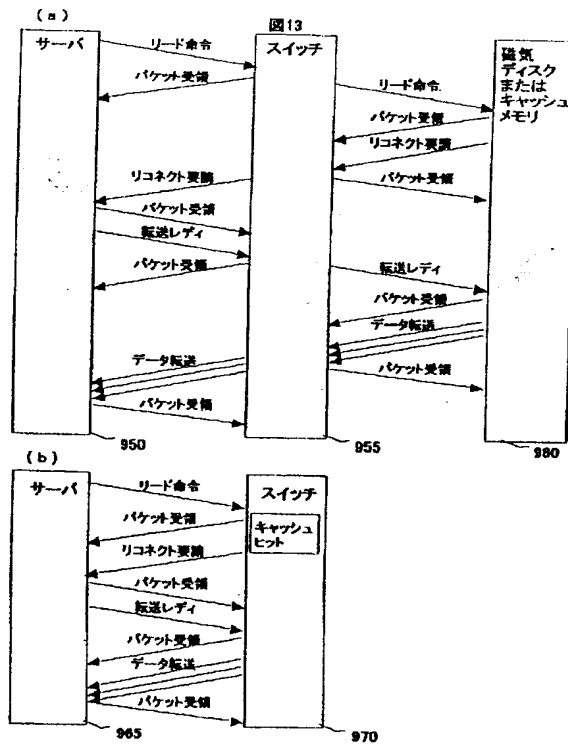


【図11】

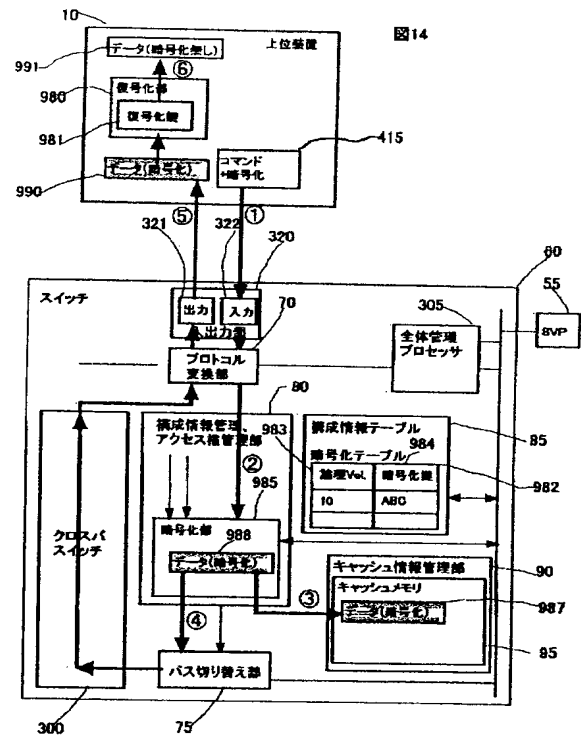
図11



【図13】



【図14】



フロントページの続き

(51) Int. Cl.⁷

識別記号

F I

テーマコード(参考)

G 0 6 F 12/08
15/16
H 0 4 L 12/56

3 2 0
6 4 5

G 0 6 F 12/08
15/16
H 0 4 L 11/20

3 2 0
6 4 5
1 0 2 Z

F ターム(参考) 5B005 JJ01 MM01 NN12 NN25 UU31
5B014 EB05 GD11 HA05 HA11
5B045 BB02 BB12 BB15 BB28 BB29
BB47 DD01 DD12
5B065 BA01 BA07 CA11 CE11 CE26
CH01
5K030 GA01 HA08 HC14 HD06 HD09
KA03 KA13 KX06 KX11 KX27
LA08 LB08 LB13 MA14 MD08